

## **Chapter 1**

### ***Introduction***

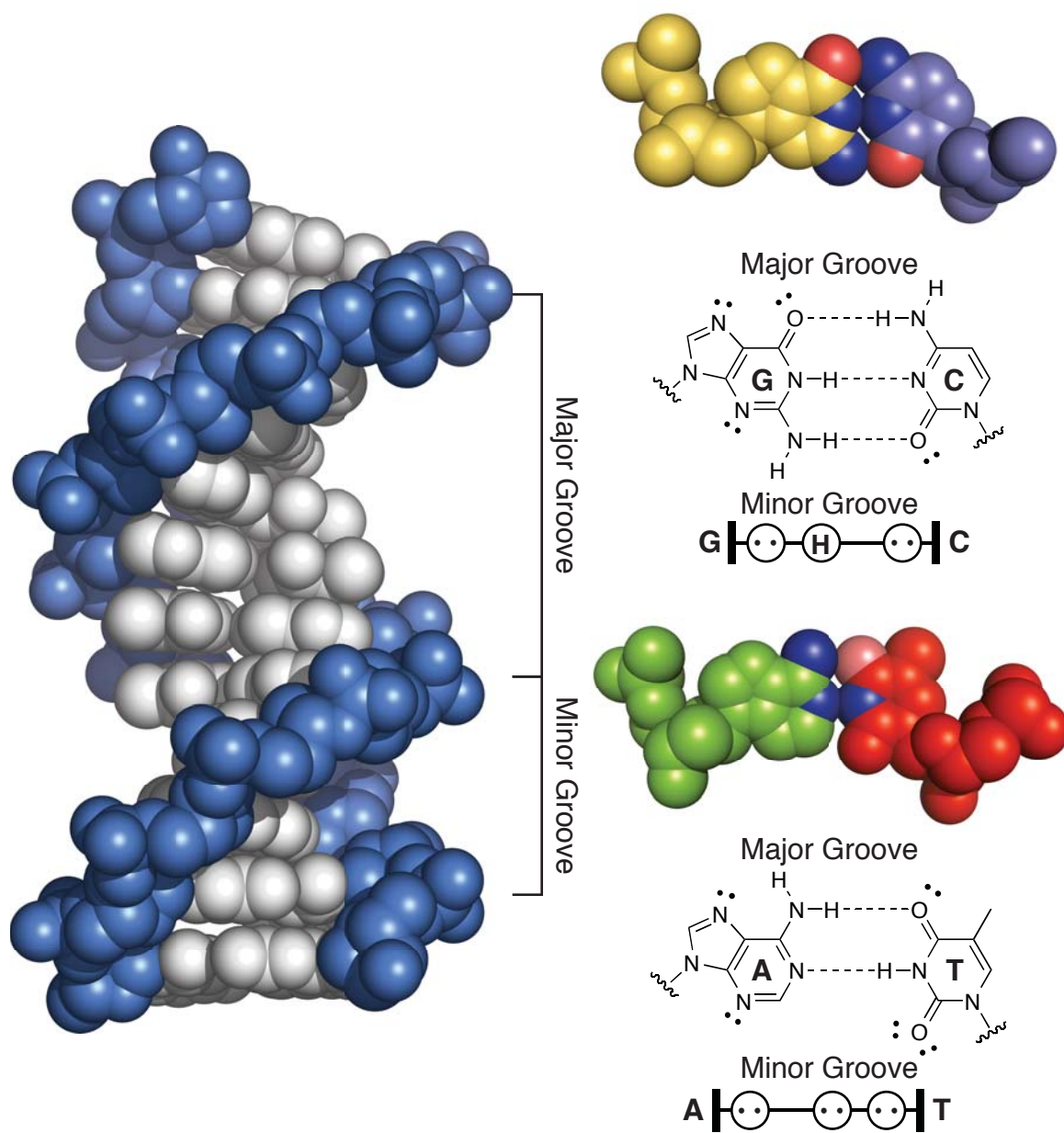
## **Deoxyribonucleic Acid (DNA)**

Deoxyribonucleic acid is the double-stranded, double-helical biopolymer that encodes all heritable information necessary to maintain life and the complex networks of gene expression and protein expression that ultimately result from its sequence. All combinations of DNA sequence polymers can be made from four monomers called nucleotides: deoxyadenosine (A), deoxycytidine (C), deoxyguanosine (G), and deoxythymidine (T), each connected to one another on a phosphodiester-linked deoxyribose sugar backbone. For each double-stranded duplex, there are rules that govern pairing of each of the four nucleotides: A pairs with T, and T with A; C pairs with G, and G with C. Several sequence and context dependent forms of DNA exist, such as A-form, B-form, and Z-form.<sup>1,2</sup> Of these, I will focus the discussion on B-form DNA as it will be most relevant to this thesis.

An X-ray crystal structure of B-form DNA is shown in Figure 1.1.<sup>1</sup> The three-dimensional structure of the clefts within DNA can be subdivided into a “major” groove and a “minor” groove. The phosphodiester-linked deoxyribose sugar backbone of DNA is highlighted in blue, and the individual bases are shown in a silver color. Each base pair is oriented perpendicularly to the helical axis of DNA. Ten base pairs of DNA represent a full turn of the DNA helix. Space-filling representations of A·T and G·C base pairs are shown in Figure 1.1. Two hydrogen bonds form between A·T base pairs and three between G·C base pairs. Beneath each base pair is shown the minor groove hydrogen bonding pattern.

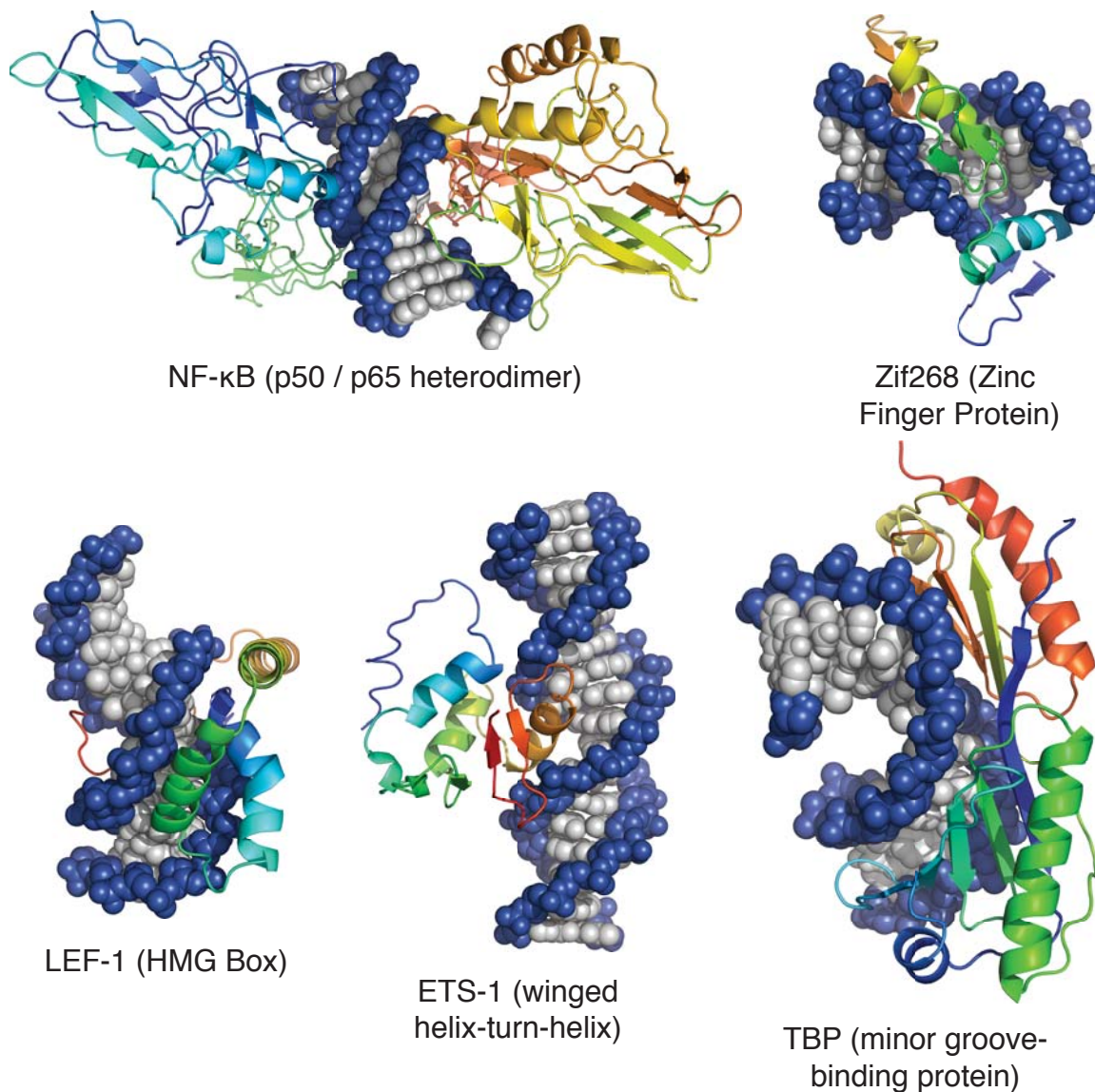
## **Protein-DNA Recognition**

The three billion base pairs of DNA within a human cell nucleus contains the information which encodes 20,000 to 25,000 protein-coding genes.<sup>3</sup> DNA is transcribed into mRNA, which forms the template for the synthesis of proteins, a process called translation. Some of the proteins produced function as transcription factors which regulate the expression of genes at the DNA→mRNA level. These proteins recognize and bind specific sequences of DNA. An example of five such proteins is denoted in Figure 1.2.



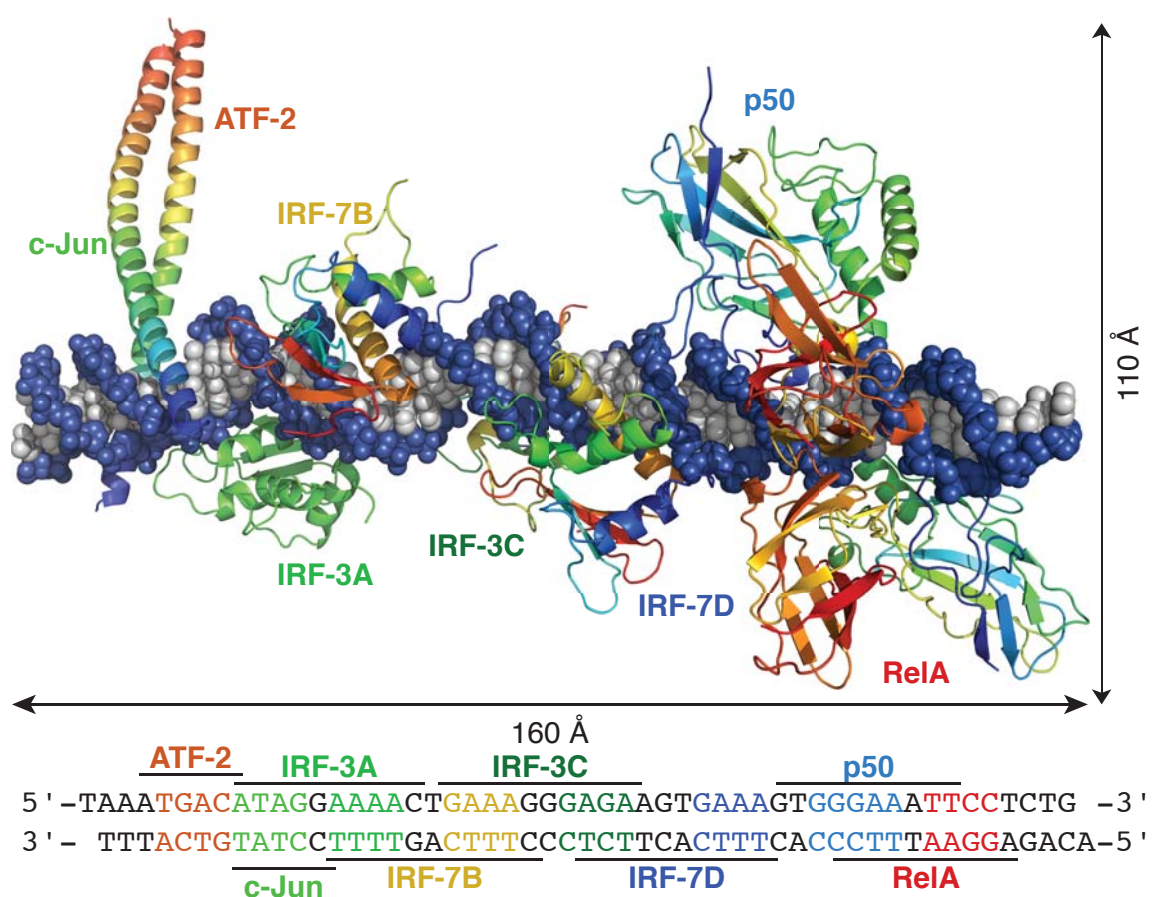
**Figure 1.1.** X-ray crystal structure (PDB 1BNA) of double-stranded DNA and examination of individual DNA base pairs

Transcription factors can be segmented into three recognition groups: a group that targets the minor groove of DNA, a group that binds the major groove, and a group that binds a hybrid of the two grooves. Of the proteins shown in Figure 1.2, NF- $\kappa$ B (p50/p65 heterodimer), Zif268, and Ets-1 recognize the major groove of DNA.<sup>4-6</sup> TATA-Binding Protein (TBP) and LEF-1 each recognize the minor groove of DNA.<sup>7,8</sup> Of these proteins, zinc finger proteins (such as Zif268) have been evolved to target new sequences of 9–18 base pairs.<sup>9</sup>



**Figure 1.2.** X-ray crystal and NMR structures of five protein–DNA complexes: NF- $\kappa$ B–DNA (PDB 1VKX), Zif268–DNA (PDB 1ZAA), LEF-1–DNA (PDB 2LEF), ETS-1–DNA (PDB 2STW), and TBP–DNA (PDB 1TGH)

Specificities of *in vitro* binding preferences for transcription factors have been studied by selective mutation studies to known binding sites and by SELEX (Systematic Evolution of Ligands by Exponential Enrichment), which gives rise to a preferred match site for the protein.<sup>10–13</sup> These *in vitro* binding data have been cataloged in both the TRANSFAC database and the JASPAR database.<sup>14,15</sup> More recent work has examined protein-DNA specificity by using fluorophore-labeled (using both covalent and non-covalent labeling schemes) proteins in conjunction with custom DNA microarrays to rank-order sequence preferences and extract sequence motifs.<sup>16–19</sup>



**Figure 1.3.** The interferon beta enhanceosome: a composite model of allosterically-driven protein–DNA recognition created from overlaid X-ray crystal structures (PDB 2O6G, 2O6I)

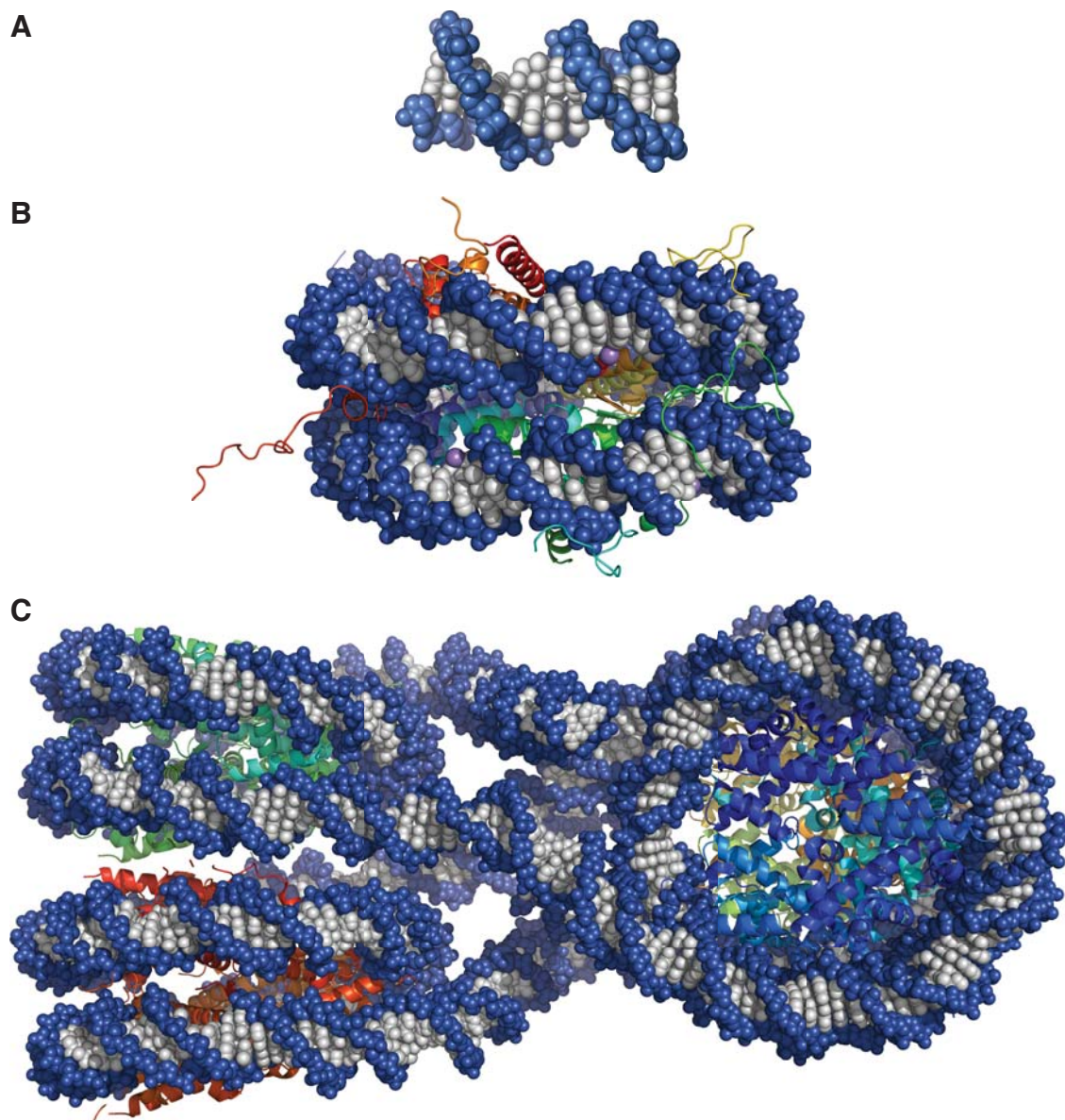
## **Allosteric Recognition of DNA**

The first example of allosterically modulated protein-DNA specificity on an enhancer was recently observed by Harrison, Maniatis, and co-workers, who used two crystal structures of proteins complexed with DNA (ATF-2/c-Jun, IRF-3A, IRF-7B, IRF-3C, IRF-7D, p50, and RelA) to create a model structure of the interferon-beta enhanceosome (Figure 1.3).<sup>20</sup> There are no protein-protein contacts between any of the proteins on this DNA sequence. Thus, it has been proposed that structural alterations to the DNA, such as widening or narrowing of the major or minor groove by individual protein-DNA interactions creates optimum binding shape and structure for other proteins, in a cooperative interaction.<sup>20</sup>

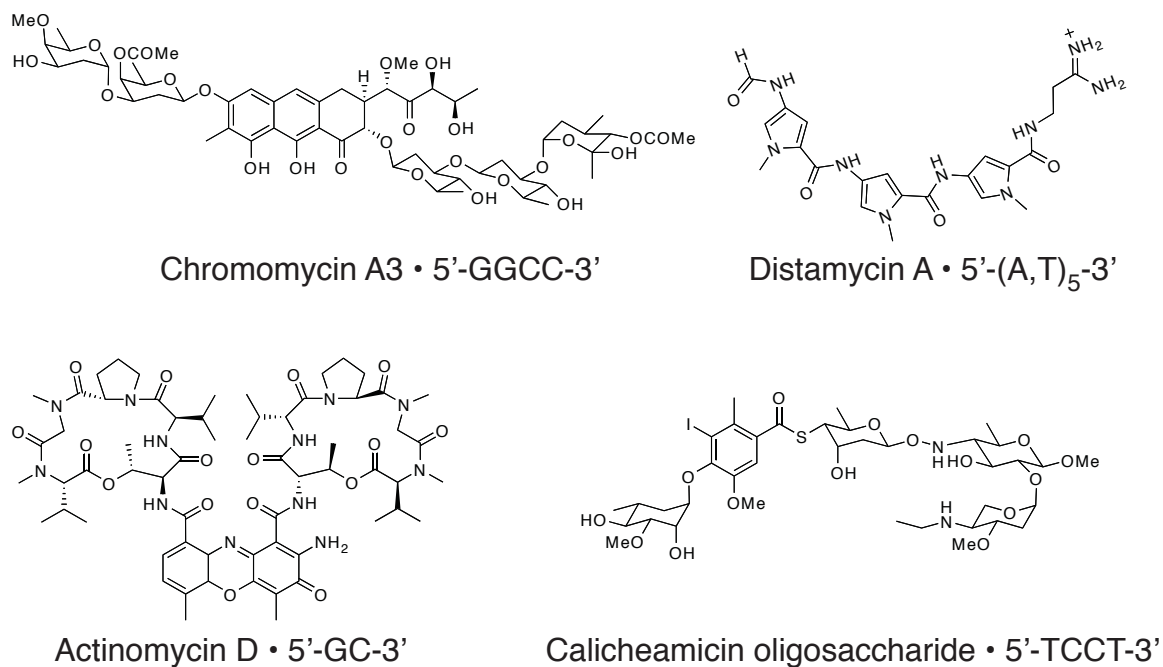
## **Higher Order DNA Structures**

While discussion of DNA to this point has examined its structure in the absence and presence of proteins, the structure found inside cellular environments is far more complex. To demonstrate the layers of folding that occur as DNA is compacted inside of a cell nucleus, Figure 1.4 shows the progression from naked DNA (Figure 1.4A)<sup>1</sup> to the nucleosome (Figure 1.4B)<sup>21</sup> to the tetranucleosome (Figure 1.4C).<sup>22</sup> These are the known high-resolution structures. Packing continues to chromatin fibre, which has been modeled,<sup>22</sup> and with more structural proteins, eventually to a single chromosome. The transition of DNA to the nucleosome is facilitated by an histone octamer, consisting of two copies each of histones H2A, H2B, H3, and H4. Duplex DNA wraps twice about the histone octamer, forming a 146–147 bp nucleosome.<sup>21</sup> Until recently, this was the only high-resolution structural data known about the repeating unit in chromatin packing. An X-ray crystal structure of a tetranucleosome solved at 9 Å resolution answered the long-standing question of how four nucleosomes would pack together.<sup>22</sup>





**Figure 1.4.** Known X-ray Crystal Structures Demonstrating Higher-order DNA Packing. a) X-ray crystal structure of double-stranded DNA (PDB 1BNA). b) X-ray crystal structure of 147 bp nucleosome core particle (PDB 1KX5). c) X-ray crystal structure of tetranucleosome, showing how nucleosomes pack (PDB 1ZBB)

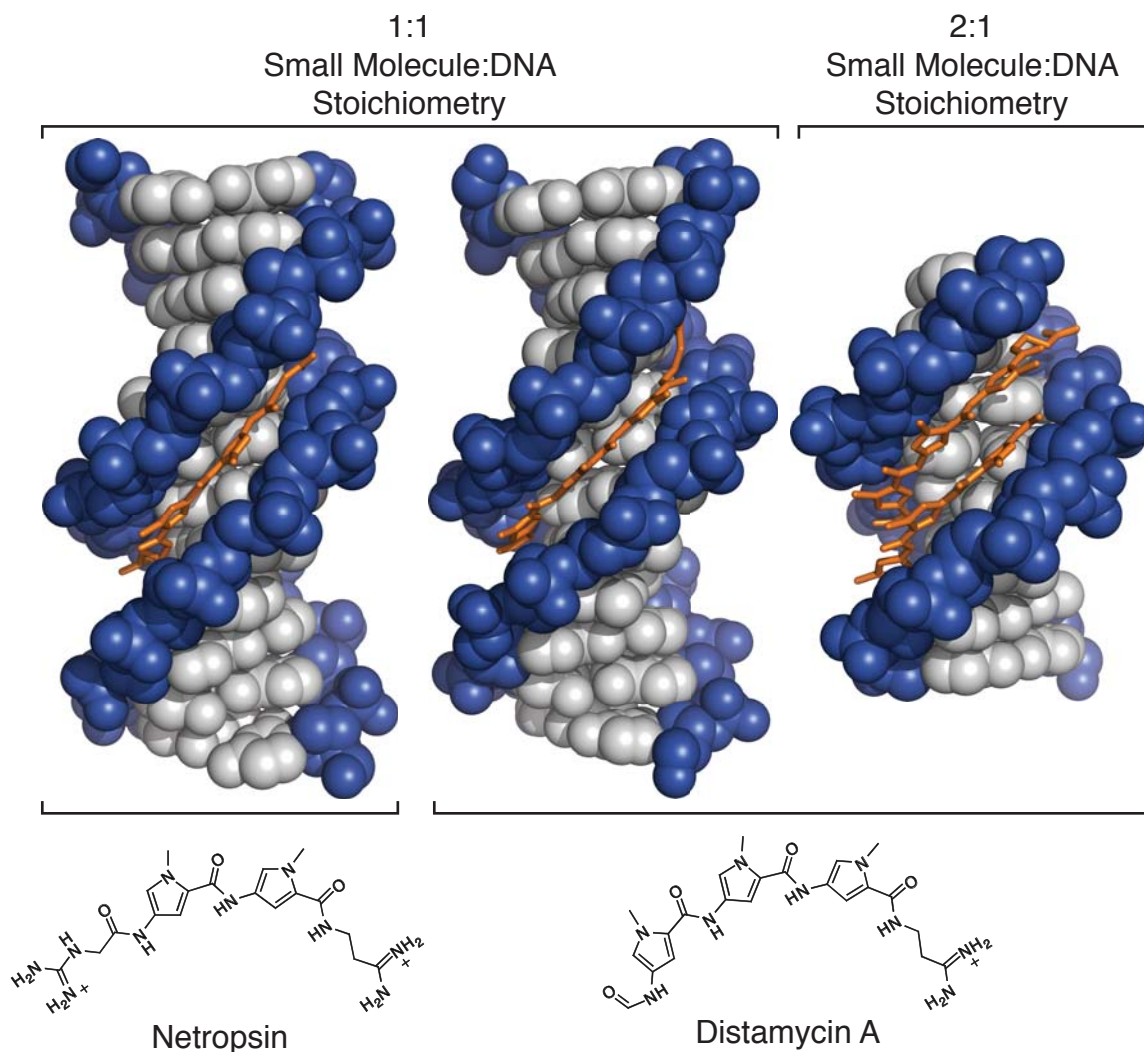


**Figure 1.5.** Natural products that recognize DNA: chromomycin, distamycin, actinomycin D, and calicheamicin oligosaccharide

### Natural Products that Recognize DNA

In addition to the nuclear proteins that recognize and bind DNA, there are numerous natural products that bind DNA in a sequence-specific fashion. Four are shown in Figure 1.5: chromomycin, distamycin, actinomycin D, and calicheamicin oligosaccharide. Chromomycin recognizes the minor groove of DNA, targets the sequence 5'-GGCC-3', and binds in a 2:1 ligand:DNA stoichiometry. Chromomycin is believed to be biologically active by interfering with replication and transcription.<sup>23</sup> Actinomycin D intercalates DNA preferentially at 5'-GC-3' sequences in a 1:1 ligand:DNA stoichiometry. It has been used as a chemotherapeutic and is known to inhibit transcription and potentially DNA replication as well.<sup>24,25</sup> Calicheamicin oligosaccharide binds the DNA minor groove as a monomer and recognizes 5'-TCCT-3'.<sup>26</sup>





**Figure 1.6.** X-ray crystal structures showing the molecular recognition of DNA by netropsin (PDB 6BNA) and distamycin (PDB 2DND in a 1:1 complex with DNA and PDB 378D in a 2:1 complex with DNA), both oligopyrrole natural products

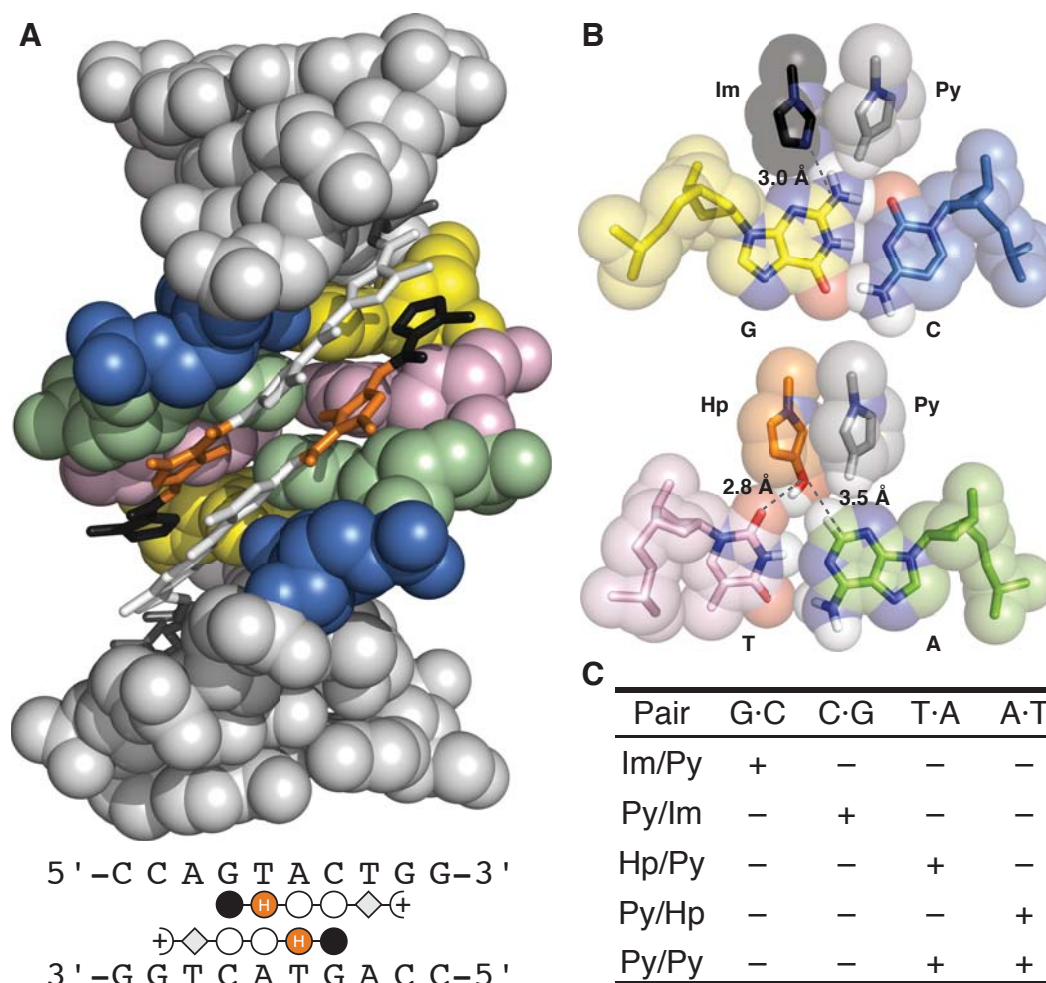
### Molecular Recognition of DNA by Distamycin and Netropsin

Distamycin and netropsin are both low molecular weight oligomers of *N*-methylpyrrole carboxamides, with distamycin containing three monomer units and netropsin containing two.<sup>27,28</sup> X-ray crystal structures of netropsin and distamycin are shown in Figure 1.6.<sup>27,29–31</sup> Netropsin and distamycin each recognize A,T tracts of DNA. Netropsin is known to bind in a 1:1 ligand:DNA complex,<sup>29</sup> whereas distamycin can bind both in a 1:1 and 2:1 ligand:DNA stoichiometry.<sup>30,31</sup> The positively charged tails at the N- and C- termini of netropsin likely prevent 2:1 binding stoichiometry.

Binding in 1:1 and 2:1 ligand:DNA stoichiometries alters the width and depth of the minor groove of DNA. In the 1:1 ligand:DNA binding cases, the minor groove of DNA narrows and deepens upon ligand binding.<sup>27,29,30</sup> In the 2:1 ligand:DNA binding case, distamycin binds as an antiparallel dimer and widens the minor groove. The hydrogen bonding patterns present in the minor groove are presented more closely to the distamycin, making a shallower groove.<sup>31</sup> Through the X-ray crystal structures of these ligands, it was originally suggested that the minor groove-protruding exocyclic amine of guanine in G·C base pairs may be accommodated by utilizing *N*-methylimidazole in place of *N*-methylpyrrole.<sup>27,29</sup> This substitution<sup>32–34</sup> led to the beginnings of a pairing code for binding of synthetically modified distamycin-like molecules, which have been called polyamides.<sup>35,36</sup>

### **Polyamide Recognition of the DNA Minor Groove**

The incorporation of *N*-methylimidazole (Im) into a distamycin-like polymer: Im-Py-Py-Dp yielded not a 5'-SWW-3' (S = G,C; W = A,T) recognition sequence but rather 5'-WGWCW-3', from which it was proposed that a 2:1 ligand:DNA binding stoichiometry could account for the observation.<sup>32,33,35</sup> These side-by-side minor groove ring pairings would form the basis for recognition of all four DNA base pairs. G·C, C·G, and A·T/T·A could be targeted using *N*-methylimidazole (Im) and *N*-methylpyrrole ligands (Py) bound side-by-side. When Im was adjacent to Py (Im / Py), G·C was read. Py / Im reads C·G, and Py / Py reads A·T or T·A. The specificity problem between A·T and T·A minor groove recognition was soon resolved with the introduction of the *N*-methyl-3-hydroxypyrrole monomer (Hp). In this case, Hp / Py reads T·A, while Py / Hp reads A·T.<sup>37</sup> When lengthening oligomers to read more base pairs, it was hypothesized that the polyamides were overcurved relative to the minor groove of DNA. Beta-alanine ( $\beta$ ) was introduced as a substitute for Py that enables the curvature of the polyamide to be reset to better align with longer sequences of DNA, while maintaining the same recognition properties as

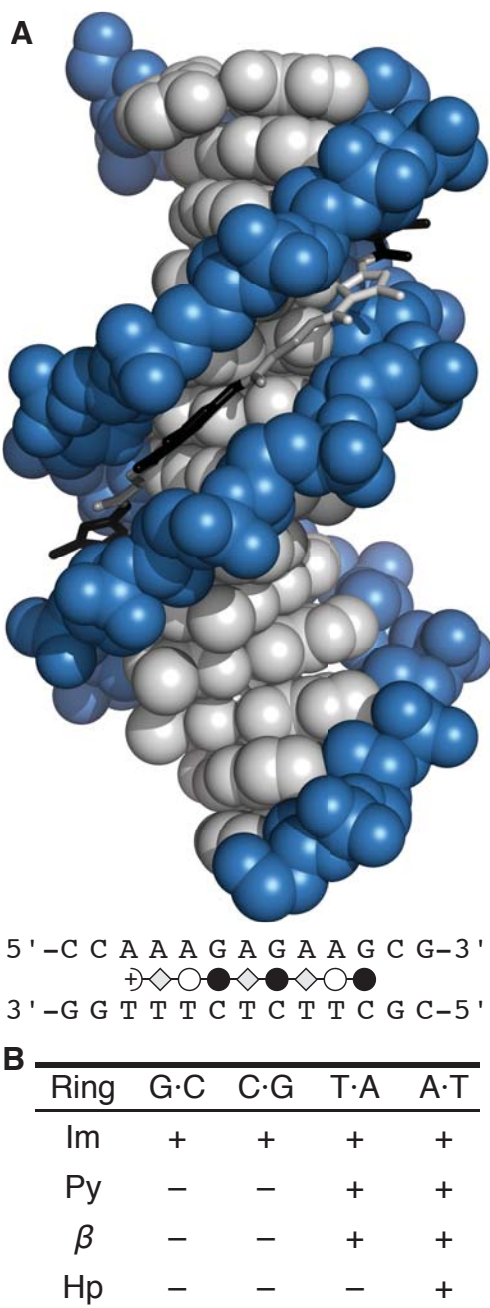


**Figure 1.7.** Recognition of all four base pairs of DNA by designed ligands based on the natural product distamycin. a) X-ray crystal structure of a 2:1 ligand:DNA decamer complex bound by the polyamide ligand Im-Hp-Py-Py- $\beta$ -Dp (PDB 407D). Filled circles represent Im; orange circles with white H, Hp; hollow circles, Py; greyed diamonds,  $\beta$ ; and half-circle with plus sign, Dp. b) Putative hydrogen-bond contacts made between polyamide monomers and individual base pairs. c) Polyamide pairing code for minor groove recognition of duplex DNA. A “+” denotes a favored interaction, and a “—” denotes a disfavored interaction.

Py.<sup>38,39</sup> Each of these monomers is shown in Figure 1.9B.

X-ray crystallography enabled the full structural understanding of DNA minor-groove recognition by polyamides, first for the basis for the G·C specific recognition<sup>40</sup> and then for A·T specific recognition (Figure 1.7).<sup>41</sup> The 2:1 binding polyamides generally orient themselves from N-terminus to C-terminus relative to DNA read 5' to 3'.<sup>42</sup>

While the 2:1 binding stoichiometry enabled all four base pairs to be read, linear beta-

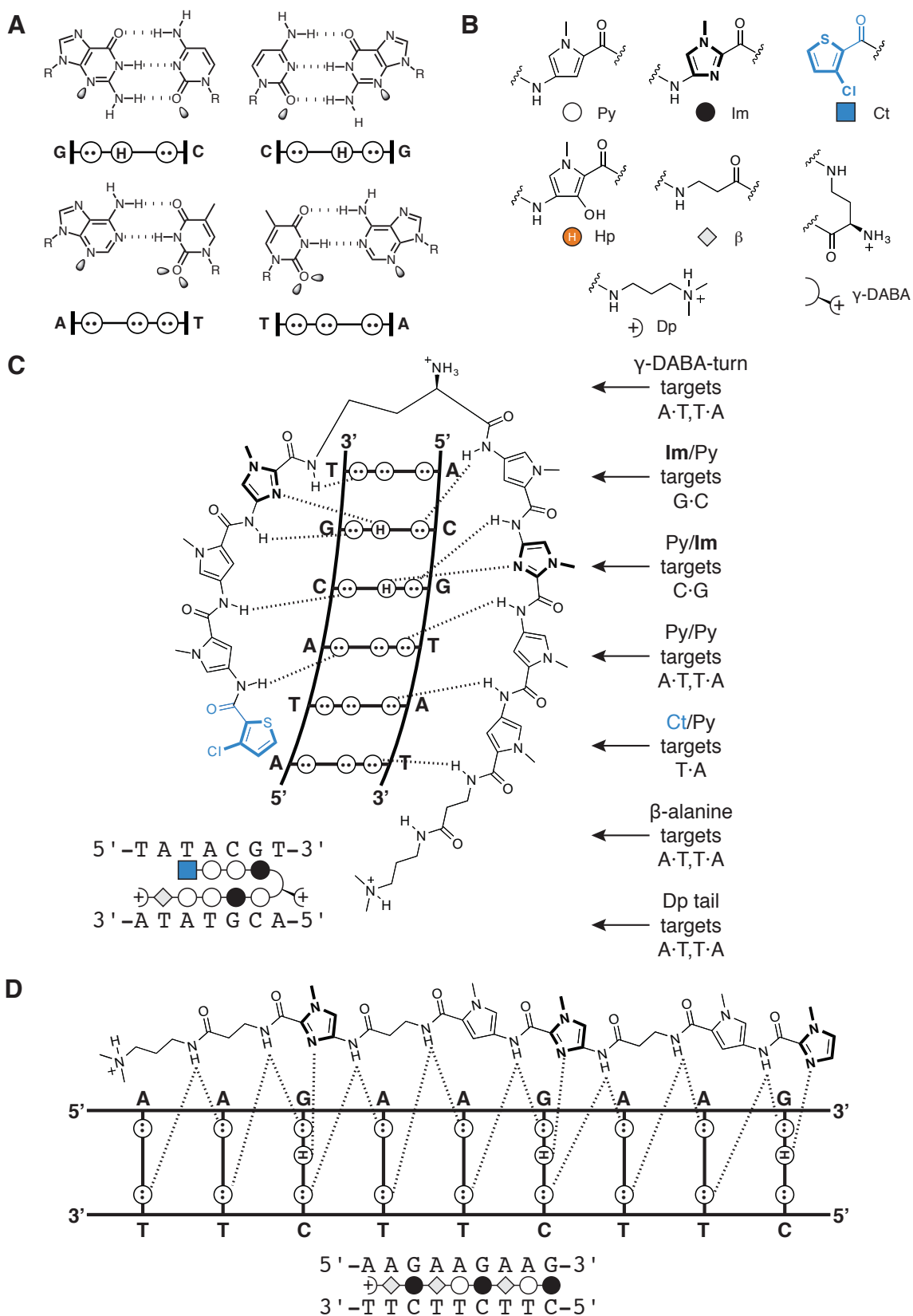


**Figure 1.8.** Molecular recognition of narrow, minor groove DNA tracts by linear beta-alanine linked polyamides. a) Average NMR structure of Im-Py- $\beta$ -Im- $\beta$ -Im-Py- $\beta$ -Dp (PDB 1LEJ). In the ball-and-stick polyamide structure, filled circles represent Im, hollow circles Py, greyed diamonds  $\beta$ , and half-circle with a plus sign, Dp. b) Observed specificity of polyamide monomers in a 1:1 polyamide:DNA complex. A “+” indicates a favored interaction, and a “-” indicates a disfavored interaction.

alanine linked ( $\beta$ -linked) polyamides enable the targeting of narrow minor groove tracts, typically purine-rich strands of DNA, such as 5'-AAAGAGAAG-3' in a 1:1 ligand:DNA stoichiometry (Figure 1.8).<sup>43-45</sup> Im, Py,  $\beta$ , and Hp have been studied in the linear  $\beta$ -linked

**Figure 1.9.** DNA minor groove hydrogen bond recognition elements. a) Minor groove hydrogen bond donors and acceptors for all four base pairs. b) Monomers found in polyamides. The ball-and-stick symbols as well as abbreviations are listed below the chemical structures. c) Potential hydrogen bonds formed between 5'-WTACGW-3' and a hairpin polyamide, Ct-Py-Py-Im-(*R*)<sup>H<sub>2</sub>N</sup> $\gamma$ -Py-Im-Py-Py- $\beta$ -Dp. d) Potential hydrogen bonds formed between 5'-AAGAAGAAG-3' and a linear  $\beta$ -linked polyamide, Im-Py- $\beta$ -Im-Py- $\beta$ -Im- $\beta$ -Dp
































polyamide class—Im is degenerate for all four base pairs, Py and  $\beta$  prefer W (A,T) over S (G,C), and Hp prefers A·T in the homo-purine strand, although at lower binding affinity.<sup>44</sup> The linear  $\beta$ -linked polyamides orient themselves from N-terminus to C-terminus relative to DNA read from 3' to 5' on the purine-rich strand of DNA.<sup>43–45</sup> The NMR structure of a linear  $\beta$ -linked polyamide in complex with DNA has been solved.<sup>45</sup> Figure 1.9D shows an example of the hydrogen bond recognition between a polypurine tract of DNA and a linear  $\beta$ -linked polyamide that targets 5'-AAGAAGAAG-3.'

Polyamides that are currently used are generally derived from 2:1 binding class and less frequently from the 1:1 binding class. To prepay the entropic costs of polyamide dimerization, a hairpin “turn-unit,”  $\gamma$ -aminobutyric acid, was used to covalently attach the two polyamide oligomers in an antiparallel fashion (Figure 1.9C).<sup>46</sup> This covalent modification allowed subnanomolar concentrations of ligands to target 5'-WGTACW-3.<sup>47</sup> Other covalent linkages have been utilized: the H-pin polyamide motif,<sup>48</sup> the U-pin polyamide motif,<sup>49</sup> and the cyclic polyamide motif,<sup>50</sup> although the hairpin motif has been most widely studied. Because hairpin polyamides occasionally align N-terminus to C-terminus against DNA read 3' to 5,' the  $\gamma$ -aminobutyric acid linker was modified to an (*R*)-2,4-diaminobutyric acid, which increased binding affinity and enforced an N to C terminus alignment with DNA read 5' to 3'.<sup>42,51</sup> Hairpin polyamides typically are constructed from eight aromatic heterocycles, or “rings,” and target six base pairs.

Recent work has completed a library of all possible eight-ring hairpin polyamides that have an N-terminal *N*-methylimidazole monomer.<sup>52</sup> Much of the scope of known sequences that hairpin polyamides can target is shown in Table 1.1.<sup>52</sup>

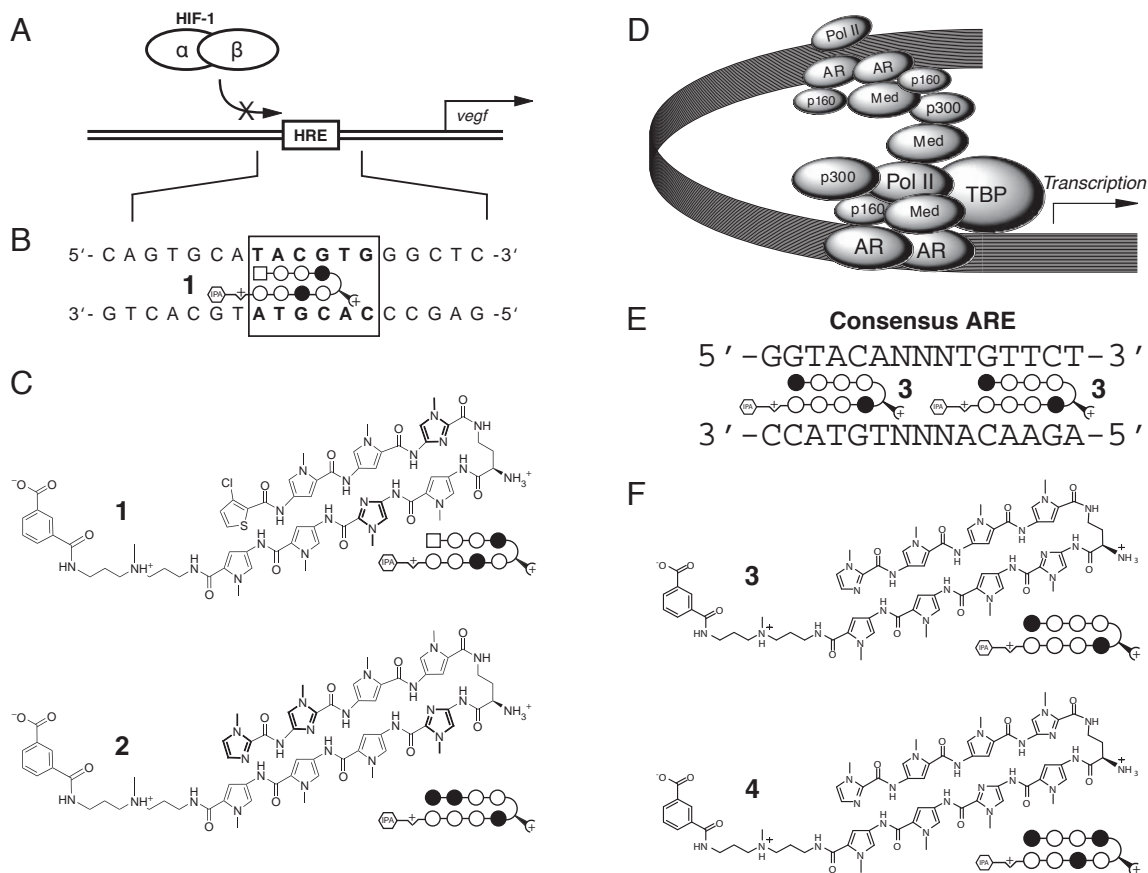
**Table 1.1.** Imidazole-capped eight-ring polyamides recognize a diverse library of DNA sequences

General (5'→3')	Polyamide	$K_a$ ( $M^{-1}$ )	Sequence context
1 WWGWWWW		$3 \times 10^9$	5'-TAGTATT-3'
2 WWGGWWW		$5 \times 10^8$	5'-CTGGTTA-3'
3 WWGWGWW		$4 \times 10^9$	5'-TAGTGAA-3'
4 WWGWGWG		$9 \times 10^9$	5'-TAGTAGT-3'
5 WWGWWCW		$3 \times 10^{10}$	5'-TAGTACT-3'
6 WWGWCWW		$2 \times 10^9$	5'-GAGTCTA-3'
7 WWGCWWW		$5 \times 10^9$	5'-ATGCAAA-3'
8 WWGGGWW		$3 \times 10^8$	5'-AAGGGAA-3'
9 WWGGWGW		$1 \times 10^{10}$	5'-TAGGTGT-3'
10 WWGGWCW		$1 \times 10^{10}$	5'-ATGGTCA-3'
11 WWGGCWW		$4 \times 10^8$	5'-AAGGCAT-3'
12 WWGWGGW		$4 \times 10^{10}$	5'-TAGTGGT-3'
13 WWGWGCW		$2 \times 10^9$	5'-ATGAGCT-3'
14 WWGCGWW		$2 \times 10^9$	5'-ATGCGTA-3'
15 WWGCWGW		$2 \times 10^9$	5'-TAGCAGT-3'
16 WWGCWCW		$9 \times 10^9$	5'-ATGCTCA-3'
17 WWGWCGW		$1 \times 10^{10}$	5'-ATGACGT-3'
18 WWGWCCW		$2 \times 10^9$	5'-TAGACCA-3'
19 WWGCCWW		$7 \times 10^8$	5'-ATGCCTA-3'
20 WWGGGGW		$2 \times 10^8$	5'-GAGGGGT-3'
21 WWGCGGW		$9 \times 10^8$	5'-ATGCGGT-3'
22 WWGGCGW		$2 \times 10^8$	5'-CAGGCGT-3'
23 WWGGGCW		$1 \times 10^8$	5'-CTGGGCA-3'
24 WWGCCGW		$2 \times 10^9$	5'-ATGCCGT-3'
25 WWGGCCW		$9 \times 10^9$	5'-ATGGCCA-3'
26 WWGCGCW		$3 \times 10^9$	5'-ATGCGCA-3'
27 WWGCCCW		$1 \times 10^9$	5'-ATGCCCA-3'

## Polyamides for Control of Gene Expression

The selective modulation of gene networks by programmable oligomers that specify short DNA sequences holds promise for new approaches to molecular medicine. The DNA minor groove-binding-polyamides have been utilized in several instances to modulate gene expression in cell culture (Figure 1.10). This ability did not come without three key studies on nuclear localization of polyamides in live cell culture.<sup>53,54</sup> Hypoxia inducible factor 1 $\alpha$  (HIF-1 $\alpha$ ) is a transcription factor that drives the expression of vascular endothelial growth factor (VEGF), a gene responsible for the vascularization of tumors. HIF-1 $\alpha$  targets the Hypoxic Response Element (HRE) consensus sequence, 5'-TACGTG-3', a subset of which is recognized by Ct-Py-Py-Im-(R)-<sup>H2N</sup> $\gamma$ -Py-Im-Py-Py-Dp-IPA (**1**). HeLa and U251 cells have been dosed with **1** and induced with deferoxamine (DFO). HIF-1 $\alpha$ -modulated genes have much lower inductions in the presence of polyamide **1**. In the presence of a polyamide that does not bind the HRE (**2**), a statistically insignificant change in VEGF expression occurs. Microarray experiments have shown that polyamide **1** downregulates a subset of genes upregulated by DFO induction. Furthermore, chromatin immunoprecipitation (ChIP) experiments demonstrated a reduced occupancy of HIF-1 $\alpha$  at the VEGF HRE in the presence of polyamide **1**, suggesting that the polyamide acts in a sequence programmed, specific manner.<sup>55-57</sup>

Gene expression regulated by androgen receptor (AR) is critical in the development and progression of prostate cancer. Prostate specific antigen (PSA) is a well-studied marker gene that correlates with the presence of prostate cancer. Androgen receptor binds as a homodimer to the androgen response element (ARE), 5'-GGTACAnnnTGTCT-3', a sequence which can readily be targeted by an eight-ring Py/Im hairpin polyamide, Im-Py-Py-Py-(R)-<sup>H2N</sup> $\gamma$ -Im-Py-Py-Py-Dp-IPA (**3**). In the presence of dihydrotestosterone (DHT) induction, LNCaP cells dosed with polyamide **3** exhibit suppressed PSA induction. A mismatch control polyamide targeting 5'-WGWCGW-3' (Im-Py-Py-Im-(R)-<sup>H2N</sup> $\gamma$ -Py-Im-Py-Py-Dp-IPA, **4**) has a much smaller effect on mRNA transcript levels of PSA. Microarray



**Figure 1.10.** Examples of polyamides used for modulation of gene expression in living cell culture. a) Structure of VEGF promoter showing inhibition of HRE binding by the HIF-1 $\alpha$ /HIF-1 $\beta$  dimer inhibits gene transcription. b) HRE enhancer binding sequence shown with match polyamide **1**. c) Match (**1**) and mismatch (**2**) polyamides utilized to target the HRE. d) Model of the androgen receptor transcription complex. e) Consensus ARE targeted by match polyamide **3**. f) Match (**3**) and mismatch (**4**) polyamides utilized to target the ARE.

experiments on **3** and **4** showed **3** to disrupt induction by DHT for a subset of the DHT-induced genes. ChIP experiments on **3** and **4** again suggest the disruption of a protein-DNA interface as a potential mechanism for polyamide activity.<sup>58</sup>

### Scope of this Thesis

In Chapter 2, we explore the use of a linear  $\beta$ -linked polyamide for upregulation of frataxin expression in cell culture. We also explore binding properties of other linear  $\beta$ -linked polyamides by DNase I footprint titrations (also in Appendix B).<sup>59</sup> In Chapter 3, we utilize DNA microarrays to examine the binding preferences of the linear  $\beta$ -linked



polyamide that modulated frataxin expression in Chapter 2. We also examine the binding preferences of the hairpin polyamide utilized to disrupt HIF-1 $\alpha$  binding (**1**). Because the microarrays necessitate a fluorophore label, we study the ramifications of different labeling positions, terminal and internal. We find that the polyamide core, and not the label position, drives DNA binding specificity of the polyamide (Appendix C). Finally, in Chapter 4, we move from *in vitro* preferences of synthetic ligands to *in vivo* preferences of protein transcription factors. We study androgen receptor binding preferences in a living cell (LNCaP cells) using ChIP-Seq, chromatin immunoprecipitation followed by high-throughput DNA sequencing. We map the genome-wide binding landscape of androgen receptor. This marks a first step toward creating a negative occupancy map of genomic regions polyamide **3** prevents androgen receptor binding.

## References

- (1) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 2179–83.
- (2) Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V.; Kopka, M. L. *Science* **1982**, *216*, 475–85.
- (3) Consortium, I. H. G. S. *Nature* **2004**, *431*, 931–45.
- (4) Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809–17.
- (5) Werner, M. H.; Clore, G. M.; Fisher, C. L.; Fisher, R. J.; Trinh, L.; Shiloach, J.; Gronenborn, A. M. *J. Biomol. NMR* **1997**, *10*, 317–28.
- (6) Chen, F. E.; Huang, D. B.; Chen, Y. Q.; Ghosh, G. *Nature* **1998**, *391*, 410–3.
- (7) Love, J. J.; Li, X.; Case, D. A.; Giese, K.; Grosschedl, R.; Wright, P. E. *Nature* **1995**, *376*, 791–5.
- (8) Juo, Z. S.; Chiu, T. K.; Leiberman, P. M.; Baikalov, I.; Berk, A. J.; Dickerson, R. E. *J. Mol. Biol.* **1996**, *261*, 239–54.
- (9) Jamieson, A. C.; Miller, J. C.; Pabo, C. O. *Nat. Rev. Drug Discov.* **2003**, *2*, 361–8.
- (10) Roulet, E.; Busso, S.; Camargo, A. A.; Simpson, A. J.; Mermoud, N.; Bucher, P. *Nat. Biotechnol.* **2002**, *20*, 831–5.
- (11) Klug, S. J.; Famulok, M. *Mol. Biol. Rep.* **1994**, *20*, 97–107.
- (12) Roulet, E.; Bucher, P.; Schneider, R.; Wingender, E.; Dusserre, Y.; Werner, T.; Mermoud, N. *J. Mol. Biol.* **2000**, *297*, 833–48.
- (13) Fields, D. S.; He, Y.; Al-Uzri, A. Y.; Stormo, G. D. *J. Mol. Biol.* **1997**, *271*, 178–94.
- (14) Wingender, E.; Dietze, P.; Karas, H.; Knuppel, R. *Nucleic Acids Res.* **1996**, *24*, 238–41.
- (15) Sandelin, A.; Alkema, W.; Engstrom, P.; Wasserman, W. W.; Lenhard, B. *Nucleic Acids Res.* **2004**, *32*, D91–4.
- (16) Berger, M. F.; Philippakis, A. A.; Qureshi, A. M.; He, F. S.; Estep, P. W., 3rd; Bulysk,

- M. L. *Nat. Biotechnol.* **2006**, *24*, 1429–35.
- (17) Bulyk, M. L.; Gentalen, E.; Lockhart, D. J.; Church, G. M. *Nat. Biotechnol.* **1999**, *17*, 573–7.
  - (18) Bulyk, M. L.; Huang, X.; Choo, Y.; Church, G. M. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 7158–63.
  - (19) Badis, G.; Chan, E. T.; van Bakel, H.; Pena-Castillo, L.; Tillo, D.; Tsui, K.; Carlson, C. D.; Gossett, A. J.; Hasinoff, M. J.; Warren, C. L.; Gebbia, M.; Talukder, S.; Yang, A.; Mnaimneh, S.; Terterov, D.; Coburn, D.; Li Yeo, A.; Yeo, Z. X.; Clarke, N. D.; Lieb, J. D.; Ansari, A. Z.; Nislow, C.; Hughes, T. R. *Mol. Cell* **2008**, *32*, 878–87.
  - (20) Panne, D.; Maniatis, T.; Harrison, S. C. *Cell* **2007**, *129*, 1111–23.
  - (21) Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J. *J. Mol. Biol.* **2002**, *319*, 1097–113.
  - (22) Schalch, T.; Duda, S.; Sargent, D. F.; Richmond, T. J. *Nature* **2005**, *436*, 138–41.
  - (23) Hou, M. H.; Robinson, H.; Gao, Y. G.; Wang, A. H. *Nucleic Acids Res.* **2004**, *32*, 2214–22.
  - (24) Kamitori, S.; Takusagawa, F. *J. Am. Chem. Soc.* **1994**, *116*, 4154–4165.
  - (25) Hou, M. H.; Robinson, H.; Gao, Y. G.; Wang, A. H. *Nucleic Acids Res.* **2002**, *30*, 4910–7.
  - (26) Bifulco, G.; Galeone, A.; Nicolaou, K. C.; Chazin, W. J.; Gomez-Paloma, L. *J. Am. Chem. Soc.* **1998**, *120*, 7183–7191.
  - (27) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 1376–80.
  - (28) Arcamone, F.; Penco, S.; Orezzi, P.; Nicoletta, V.; Pirelli, A. *Nature* **1964**, *203*, 1064–5.
  - (29) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553–63.
  - (30) Coll, M.; Frederick, C. A.; Wang, A. H.; Rich, A. *Proc. Natl. Acad. Sci. U. S. A.*

**1987**, 84, 8385–9.

- (31) Mitra, S. N.; Wahl, M. C.; Sundaralingam, M. *Acta Crystallogr.* **1999**, D55, 602–9.
- (32) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, 89, 7586–90.
- (33) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, 114, 8783–8794.
- (34) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, 266, 646–50.
- (35) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, 9, 2215–35.
- (36) Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, 13, 284–99.
- (37) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, 391, 468–71.
- (38) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, 118, 6160–6166.
- (39) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, 120, 6219–6226.
- (40) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Biol.* **1998**, 5, 104–9.
- (41) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, 282, 111–5.
- (42) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, 1997, 8756–8765.
- (43) Dervan, P. B.; Urbach, A. R. In *Essays in Contemporary Chemistry*; Quinkert, G., Kisakürek, M. V., Eds.; Verlag Helvetica Chimica Acta: Zurich, 2000, p 327–339.
- (44) Urbach, A. R.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98, 4343–8.
- (45) Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B. *J. Mol. Biol.* **2002**, 320, 55–71.

- (46) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983–7988.
- (47) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559–61.
- (48) Greenberg, W. A.; Baird, E. E.; Dervan, P. B. *Chem.-Eur. J.* **1998**, *4*, 796–805.
- (49) Heckel, A.; Dervan, P. B. *Chem.-Eur. J.* **2003**, *9*, 3353–66.
- (50) Cho, J.; Parks, M. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 10389–92.
- (51) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382–1391.
- (52) Hsu, C. F.; Phillips, J. W.; Trauger, J. W.; Farkas, M. E.; Belitsky, J. M.; Heckel, A.; Olenyuk, B. Z.; Puckett, J. W.; Wang, C. C.; Dervan, P. B. *Tetrahedron* **2007**, *63*, 6146–6151.
- (53) (a) Belitsky, J. M.; Leslie, S. J.; Arora, P. S.; Beerman, T. A.; Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 3313–8. (b) Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12063–8.
- (54) Edelson, B. S.; Best, T. P.; Olenyuk, B.; Nickols, N. G.; Doss, R. M.; Foister, S.; Heckel, A.; Dervan, P. B. *Nucleic Acids Res.* **2004**, *32*, 2802–18.
- (55) Olenyuk, B. Z.; Zhang, G. J.; Klco, J. M.; Nickols, N. G.; Kaelin, W. G., Jr.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16768–73.
- (56) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. *ACS Chem. Biol.* **2007**, *2*, 561–71.
- (57) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. *Nucleic Acids Res.* **2007**, *35*, 363–70.
- (58) Nickols, N. G.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10418–23.
- (59) Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450–66.